Two-Way Analysis of Variance

Food consumption of rats

Gender	Lard: Fresh	Lard: Rancid
Male	709	592
	679	538
	699	476
Female	657	508
	594	505
	677	539

Two-way vs one-way vs nested ANOVA

	T	
	1	2
	*	*
M	*	*
	*	*
	*	*
F	*	*
	*	*

T			
M1	M2	F1	F2
*	*	*	*
*	*	*	*
*	*	*	*

T			
1		2	
M	F	M	F
*	*	*	*
*	*	*	*
*	*	*	*

Two-way versus one-way ANOVA

In the lard example, we could consider the lard by gender groups as four different treatments, and carry out a standard one-way ANOVA.

Let

- r be the number of rows in the two-way ANOVA,
- c be the number of columns in the two-way ANOVA,
- n be the number of observations within each of those $r \times c$ groups.

One-way ANOVA table

source	sum of squares	df
between groups	$SS_{between} = n \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{\cdots})^2$	$rc - 1$
within groups	$SS_{within} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$rc(n - 1)$
total	$SS_{total} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\cdots})^2$	$rcn - 1$

Example

source	SS	df	MS	F	p-value
between	65904	3	21968	15.1	0.0012
within	11667	8	1458		

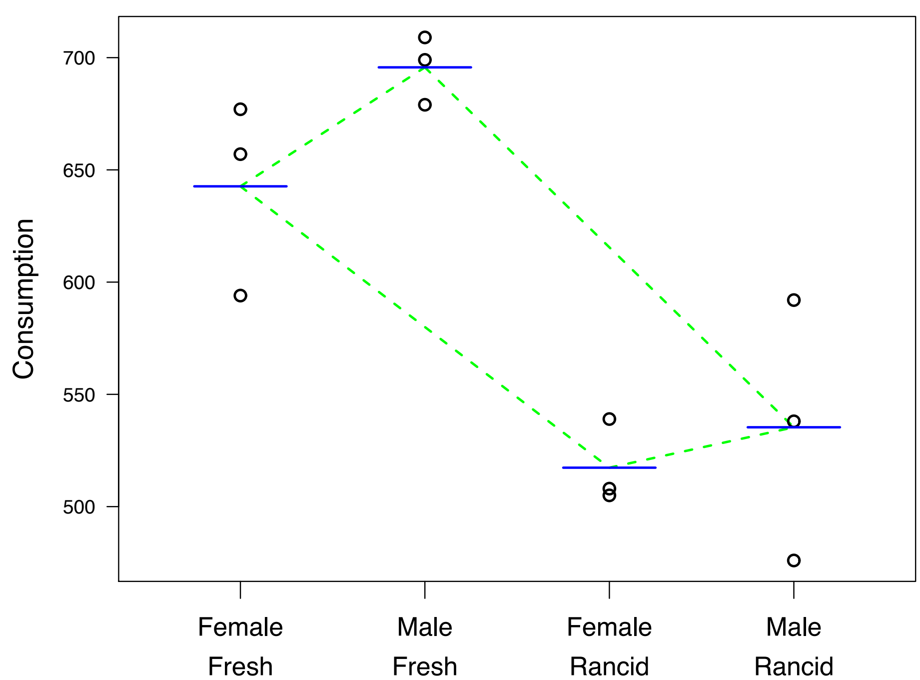
$\longrightarrow$ But this doesn't tell us anything about the separate effects of freshness and sex.

All sorts of means

	Fat		
Gender	Fresh	Rancid	
Male	695.67	535.33	615.50
Female	642.67	517.33	580.00
	669.17	526.33	597.75

⟶ This table shows the cell, row, and column means, plus the overall mean.

A picture



Two-way ANOVA table

source	sum of squares	df
between rows	$SS_{rows} = c\,n \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$	$r - 1$
between columns	$SS_{columns} = r\,n \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$c - 1$
interaction	$SS_{interaction}$	$(r - 1)(c - 1)$
error	$SS_{within} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$	$rc(n - 1)$
total	$SS_{total} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$	$rcn - 1$

Example

source	sum of squares	df	mean squares
sex	3781	1	3781
freshness	61204	1	61204
interaction	919	1	919
error	11667	8	1458

The ANOVA model

Let Y_{ijk} be the k[th] item in the subgroup representing the i[th] group of treatment A (r levels) and the j[th] group of treatment B (c levels). We write

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

The corresponding analysis of the data is

$$y_{ijk} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.})$$

Towards hypothesis testing

source	mean squares	expected mean squares
between rows	$\frac{c\,n \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2}{r-1}$	$\sigma^2 + \frac{c\,n}{r-1} \sum_i \alpha_i^2$
between columns	$\frac{r\,n \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2}{c-1}$	$\sigma^2 + \frac{r\,n}{c-1} \sum_j \beta_j^2$
interaction	$\frac{n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2}{(r-1)(c-1)}$	$\sigma^2 + \frac{n}{(r-1)(c-1)} \sum_i \sum_j \gamma_{ij}^2$
error	$\frac{\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2}{r\,c\,(n-1)}$	σ^2

This is for fixed effects, and equal number of observations per cell!

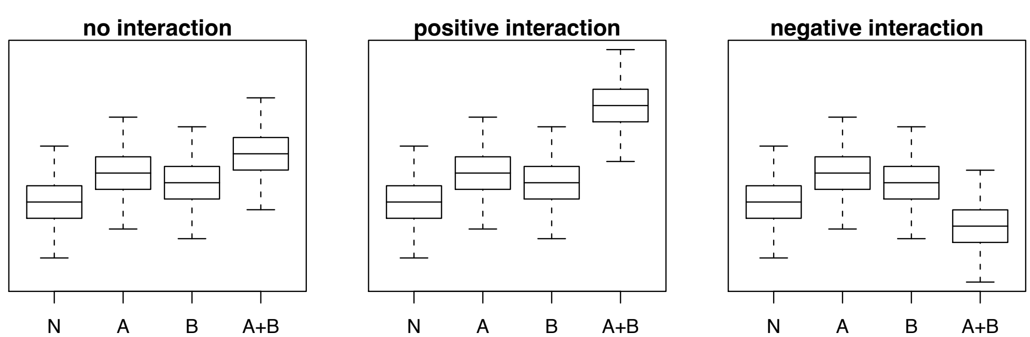
Example (continued)

source	SS	df	MS	F	p-value
Sex	3781	1	3781	2.6	0.1460
Freshness	61204	1	61204	42.0	0.0002
interaction	919	1	919	0.6	0.4503
error	11667	8	1458		

Interaction in a 2-way ANOVA model

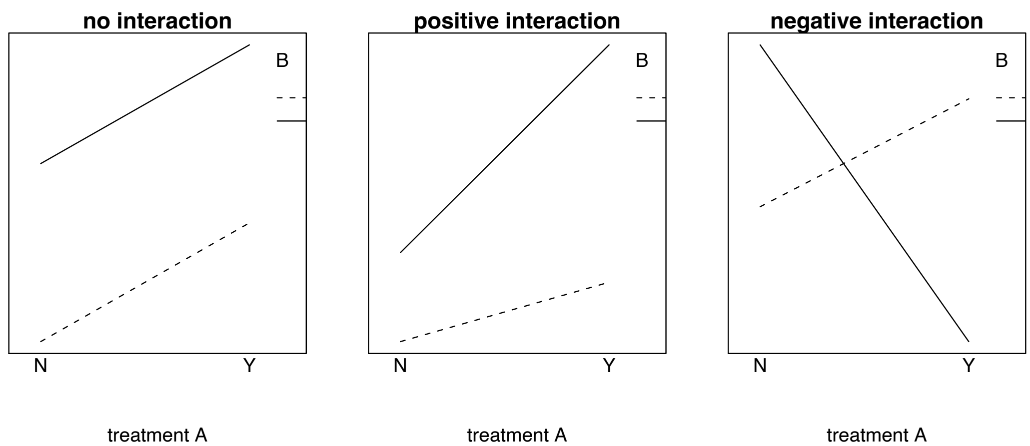
Let Y_{ijk} be the k^{th} item in the subgroup representing the i^{th} group of treatment A (r levels) and the j^{th} group of treatment B (c levels). We write

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$



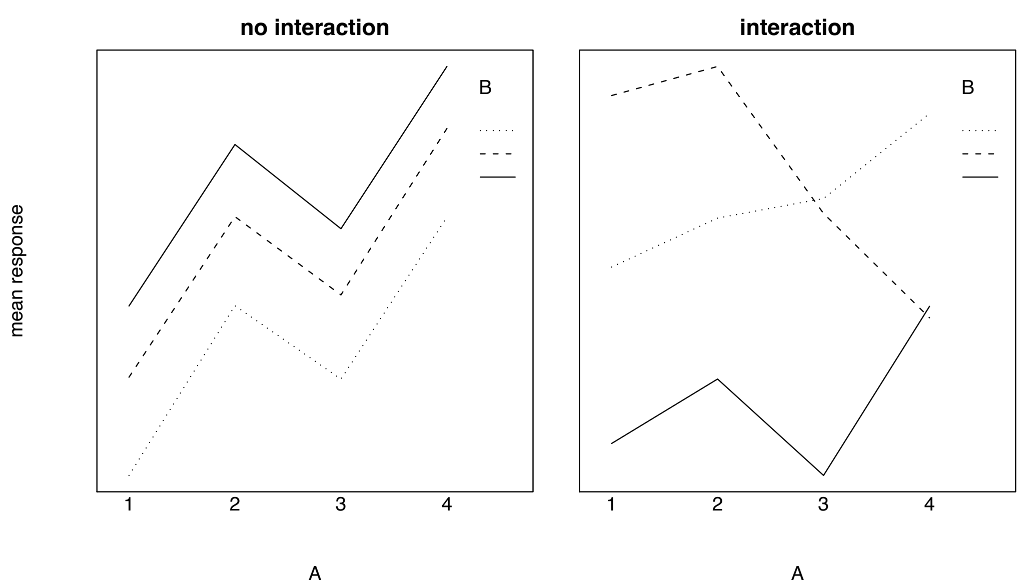
Interaction plots

The R function interaction.plot() lets you compare the cell means by treatments.



Interaction plots (2)

Assume treatment A has four levels and treatment B has three levels. The interaction plots could look like one of these:



Expected mean squares

source	mean squares	expected mean squares
between rows	$\frac{c\,n \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2}{r-1}$	$\sigma^2 + \frac{c\,n}{r-1} \sum_i \alpha_i^2$
between columns	$\frac{r\,n \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2}{c-1}$	$\sigma^2 + \frac{r\,n}{c-1} \sum_j \beta_j^2$
interaction	$\frac{n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2}{(r-1)(c-1)}$	$\sigma^2 + \frac{n}{(r-1)(c-1)} \sum_i \sum_j \gamma_{ij}^2$
error	$\frac{\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2}{r\,c\,(n-1)}$	σ^2

Expected mean squares

source	fixed effects	random effects	mixed effects
between rows	$\sigma^2 + \frac{c\,n}{r-1} \sum_i \alpha_i^2$	$\sigma^2 + n\,\sigma^2_{R\times C} + c\,n\,\sigma^2_R$	$\sigma^2 + n\,\sigma^2_{R\times C} + \frac{c\,n}{r-1} \sum_i \alpha_i^2$
between columns	$\sigma^2 + \frac{r\,n}{c-1} \sum_j \beta_j^2$	$\sigma^2 + n\,\sigma^2_{R\times C} + r\,n\,\sigma^2_C$	$\sigma^2 \qquad + r\,n\,\sigma^2_C$
interaction	$\sigma^2 + \frac{n}{(r-1)(c-1)} \sum_i \sum_j \gamma_{ij}^2$	$\sigma^2 + n\,\sigma^2_{R\times C}$	$\sigma^2 + n\,\sigma^2_{R\times C}$
error	σ^2	σ^2	σ^2

Example

Strain differences and daily differences in blood pH for five (r = 5) inbred strains of mice. Five (n = 5) mice from each strain were tested six times (c = 6) at one-week intervals.

Source	SS	df	MS	F	P-value
mouse strains	0.37	4	0.0920	17.7	< 0.001
test days	0.05	5	0.0101	1.9	0.132
interaction	0.10	20	0.0052	1.5	0.083
error	0.41	120	0.0034		

Example

mouse strains	0.0029	42.2%
test days	0.0002	2.9%
interaction	0.0004	5.3%
error	0.0034	49.6%

Two-way versus nested ANOVA revisited

	Strains				
Days	1	2	3	4	5
1	*	*	*	*	*
2	*	*	*	*	*
3	*	*	*	*	*
4	*	*	*	*	*
5	*	*	*	*	*
6	*	*	*	*	*

	Strains																													
	1						2						3						4						5					
Days	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Mice	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

ANOVA tables

	source	SS	df	MS	F
Correct:	two-way anova				
	mouse strains	0.37	4	0.0920	17.7
	test days	0.05	5	0.0101	1.9
	interaction	0.10	20	0.0052	1.5
	error	0.41	120	0.0034	
	total	0.93	149		
Incorrect:	nested anova				
	mouse strains	0.37	4	0.0920	14.8
	days within strains	0.15	25	0.0062	1.8
	error	0.41	120	0.0034	
	total	0.93	149		

Unequal number of observations

The following data were obtained in a study on energy utilization (in kcal/g) of the pocket mouse during hibernation at different temperatures.

Restricted food		Unrestricted food	
8ºC	18ºC	8ºC	18ºC
62.69	72.60	95.73	101.19
54.07	70.97	63.95	76.88
65.73	74.32	144.30	74.08
62.98	53.02	144.30	81.40
	46.22		66.58
	59.10		84.38
	61.79		118.95
	61.89		118.95

R is for rescue...

The computations for the ANOVA table get rather complicated if the numbers of observations per cell are not equal. However, you can simply use aov() to get the results.

```
> mouse.aov <- aov(log(rsp) ~ food * temp, data=mouse)
> anova(mouse.aov)
             Df  Sum Sq  Mean Sq  F value    Pr(>F)
food          1   1.07      1.07     21.4   < 0.01
temp          1   0.04      0.04      0.8     0.38
food:temp     1   0.05      0.05      1.0     0.33
Residuals    20   1.00      0.05
```

Two-way ANOVA without replicates

Below are the development periods (in days) for three strains of houseflies at seven densities.

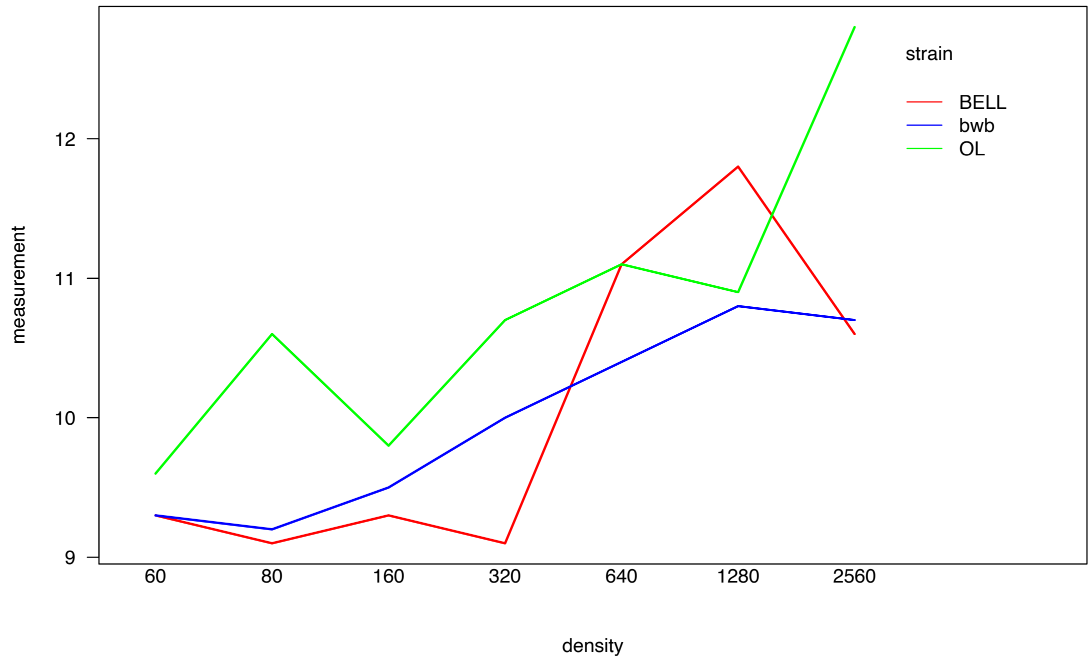
	Strain		
Density	OL	BELL	bwb
60	9.6	9.3	9.3
80	10.6	9.1	9.2
160	9.8	9.3	9.5
320	10.7	9.1	10.0
640	11.1	11.1	10.4
1280	10.9	11.8	10.8
2560	12.8	10.6	10.7

ANOVA table

source	**df**	**SS**	**MS**
fly strains	2	2.79	1.39
condition	6	12.54	2.09
interaction	12	4.11	0.34
total	20		

We have 21 observations. That means we have no degrees of freedom left to estimate an error!

Interactions



Result

If we assume that there are no interactions, i.e., we assume

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

we get the following results using aov() in R:

```
> fly.aov <- aov(rsp ~ strain + density, data=fly)
> anova(fly.aov)
             Df   Sum Sq   Mean Sq   F value    Pr(>F)
strain        2     2.79      1.39      4.07     0.045
density       6    12.54      2.09      6.10     0.004
Residuals    12     4.11      0.34
```

Expected mean squares

In general, we have:

source	fixed effects	random effects	mixed effects
between rows	$\sigma^2 + \frac{c\,n}{r-1}\sum_i \alpha_i^2$	$\sigma^2 + n\,\sigma^2_{R\times C} + c\,n\,\sigma^2_R$	$\sigma^2 + n\,\sigma^2_{R\times C} + \frac{c\,n}{r-1}\sum_i \alpha_i^2$
between columns	$\sigma^2 + \frac{r\,n}{c-1}\sum_j \beta_j^2$	$\sigma^2 + n\,\sigma^2_{R\times C} + r\,n\,\sigma^2_C$	$\sigma^2 \qquad + r\,n\,\sigma^2_C$
interaction	$\sigma^2 + \frac{n}{(r-1)(c-1)}\sum_i\sum_j \gamma_{ij}^2$	$\sigma^2 + n\,\sigma^2_{R\times C}$	$\sigma^2 + n\,\sigma^2_{R\times C}$
error	σ^2	σ^2	σ^2

Expected mean squares

If n=1 and there is no interaction, we have:

source	fixed effects	random effects	mixed effects
between rows	$\sigma^2 + \frac{c}{r-1}\sum_i \alpha_i^2$	$\sigma^2 + c\,\sigma^2_R$	$\sigma^2 + \frac{c}{r-1}\sum_i \alpha_i^2$
between columns	$\sigma^2 + \frac{r}{c-1}\sum_j \beta_j^2$	$\sigma^2 + r\,\sigma^2_C$	$\sigma^2 + r\,\sigma^2_C$
error	σ^2	σ^2	σ^2

Expected mean squares

If n=1 but there is an interaction, we have:

source	fixed effects	random effects	mixed effects
between rows	$\sigma^2 + \frac{c}{r-1}\sum_i \alpha_i^2$	$\sigma^2 + \sigma^2_{R\times C} + c\,\sigma^2_R$	$\sigma^2 + \sigma^2_{R\times C} + \frac{c}{r-1}\sum_i \alpha_i^2$
between columns	$\sigma^2 + \frac{r}{c-1}\sum_j \beta_j^2$	$\sigma^2 + \sigma^2_{R\times C} + r\,\sigma^2_C$	$\sigma^2 \qquad + r\,\sigma^2_C$
interaction	$\sigma^2 + \frac{1}{(r-1)(c-1)}\sum_i \sum_j \gamma_{ij}^2$	$\sigma^2 + \sigma^2_{R\times C}$	$\sigma^2 + \sigma^2_{R\times C}$
error	σ^2	σ^2	σ^2